Improved Mapping of Radio Sources From VLBI Data by Least-Squares Fit

E. R. Rodemich
Communications Systems Research Section

This article describes a method for producing improved mapping of radio sources from VLBI data. The method described here is more direct than existing Fourier methods, is often more accurate, and runs at least as fast. The visibility data is modeled here, as in existing methods, as a function of the unknown brightness distribution and the unknown antenna gains and phases. We want to choose these unknowns so that the resulting function values are as near as possible to the observed values. If we use the RMS deviation to measure the closeness of this fit to the observed values, we are led to the problem of minimizing a certain function of all the unknown parameters. This minimization problem cannot be solved directly, but it can be attacked by iterative methods which we show converge automatically to the minimum with no user intervention. The resulting brightness distribution will furnish the best fit to the data among all brightness distributions of given resolution.

I. Introduction

This article describes a method for producing improved mapping of radio sources from VLBI data. The use of VLBI data has led to high resolution maps of radio sources in the sky [Refs. 1-4]. The data provide values of the visibility function, which is the Fourier transform of the brightness distribution. The problem of finding the unknown brightness distribution can accordingly be expressed as the problem of finding an inverse Fourier transform. The methods currently used depend on approximate inversion methods for a Fourier transform which is known on an irregularly spaced set of points.

An additional complication is that the signal received at each antenna can have an unknown gain and phase offset, depending on conditions at this antenna as well as on atmospheric conditions. This introduces unknown multiplicative factors into the visibilities which must be eliminated before inverting the Fourier transform. Iterative methods have been developed for this which use an assumed map to recalibrate the data, get a new map from these data by Fourier inversion, then repeat the procedure starting from the new map. These iterative methods require considerable user interaction as well as computer time. They are also biased in favor of certain types of brightness distributions in the resulting map.

The method described in this article is more direct than existing Fourier methods, is often more accurate, and runs at least as fast. The visibility data are modeled here, as in existing methods, as a function of the unknown brightness distribution and the unknown antenna gains and phases. We want to choose these unknowns so that the resulting function values are as near as possible to the observed values. If we use the

RMS deviation to measure the closeness of this fit to the observed values, we are led to the problem of minimizing a certain function of all the unknown parameters. This minimization problem cannot be solved directly, but it can be attacked by iterative methods which converge automatically to the minimum with no user intervention. The resulting brightness distribution will furnish the best fit to the data among all brightness distributions of given resolution.

II. The Method

Preprocessing of the data, which we are not concerned with here, furnishes time averages of the visibility function over intervals on the order of several seconds to a minute, together with estimates of standard deviation due to noise. We get a set of values, $E_{pq}^{(n)}$, for the antenna pair p, q and the n^{th} time interval. For a given value of n, only some of the possible p, q pairs may occur, either because the source is not visible from all antennas, or because some data were lost.

Each data value, $E_{pq}^{(n)}$, is modeled by a function, $F_{pq}^{(n)}$, of the unknown brightness distribution, I_m, and the unknown gains, $A_p^{(n)}$, $A_q^{(n)}$, with noise added. The form of $F_{pq}^{(n)}$ is given in Appendix A. Using the standard deviations, $\sigma_{pq}^{(n)}$, from the preprocessing, we set up the function

$$Q = \sum_n Q_n ,$$

$$Q_n = \sum_{p,q} \left| F_{pq}^{(n)} - E_{pq}^{(n)} \right|^2 \Big/ {\sigma_{pq}^{(n)}}^2 .$$

Each $F_{pq}^{(n)}$ is linear in the I_m's and contains the product $A_p^{(n)} \overline{A_q^{(n)}}$, where the bar denotes a complex conjugate. An iterative method of minimizing Q is used which repeats step (1) followed by step (2):

(1) Minimize Q by varying the I_m's, holding the $A_p^{(n)}$'s fixed.

(2) Minimize Q by varying the $A_p^{(n)}$'s, holding the I_m's fixed.

The nature of these steps is quite different. In (1), we want to solve a system of simultaneous linear equations in a large number of unknowns. In (2), this linearity is lost, but the problem breaks up into the minimization of each Q_n separately, which only involves a small number of variables.

As this iterative procedure is carried out, the value of Q always decreases, and the unknowns approach values for which no further decrease is possible. Hence the procedure must converge.

III. Results and Conclusions

The method was applied to two sets of real data and one set of simulated data. The two real data sets are based on observations in Dec. 1982 at 6 cm. The resulting maps are shown in Figs. 1 and 2. Source III is a simulated source consisting of two point sources convolved with a circularly symmetric Gaussian distribution, with 1% noise added. The output map is shown in Fig. 3.

The computer time needed for each source on the Caltech computer "PHOBOS" was 10-15 minutes, which compares favorably with the time for conventional methods. This is possible because the new method can give good maps with a coarser grid than other methods. This is shown by Fig. 4, which displays the map of Source I derived in the usual way with a fine grid, and the effect on this map if a 32 × 32 grid is used for the final inversion and "CLEANing" ("CLEAN" is a Caltech program).

The method described here has been shown to be a practical alternative to existing methods. It can construct a map without any user intervention. The resulting map is free of the biases introduced by interpolation before inversion and by the CLEAN program.

Acknowledgment

I am indebted to Dr. T. J. Pearson of the Caltech Astronomy Department for help in providing data and access to the Campus/JPL Correlation Facility located in Robinson at Caltech.

References

1. Readhead, A. C. S., and Wilkinson, P. N., "The mapping of compact radio sources from VLBI data," *Astrophysical Journal*, 223, pp. 25–36 (1978).
2. Cotton, W. D., "A method of mapping compact structure in radio sources using VLBI observations," *Astronomical Journal*, 84, pp. 1122–1128 (1979).
3. Readhead, A. C. S., Walker, R. C., Pearson, T. J., and Cohen, M. H., "Mapping radio sources with uncalibrated visibility data," *Nature*, 285, pp. 137–140 (1980).
4. Comwell, T. J., and Wilkinson, P. N., "A new method for making maps with unstable radio interferometers," *Mon. Not. R. Astro. Soc.*, 196, pp. 1067–1086 (1981).

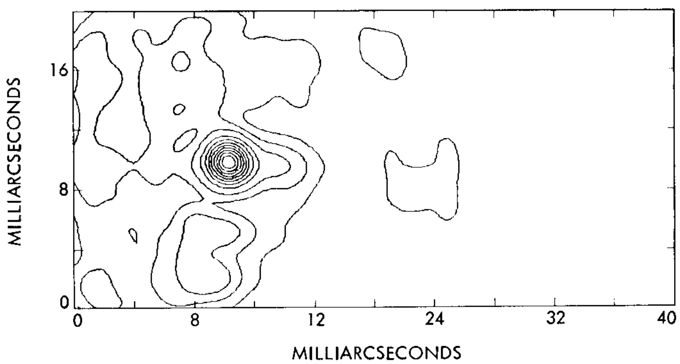

Fig. 1. Source 1807+698

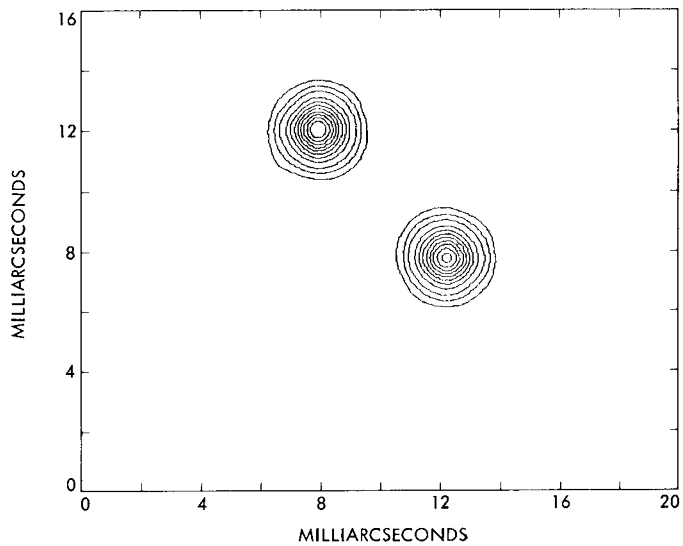

Fig. 3. Simulated source

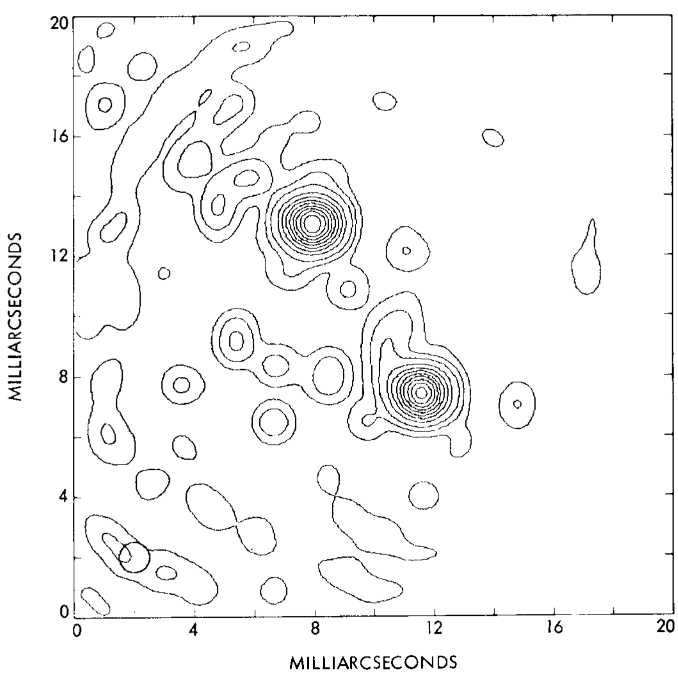

Fig. 2. Source 2021+614

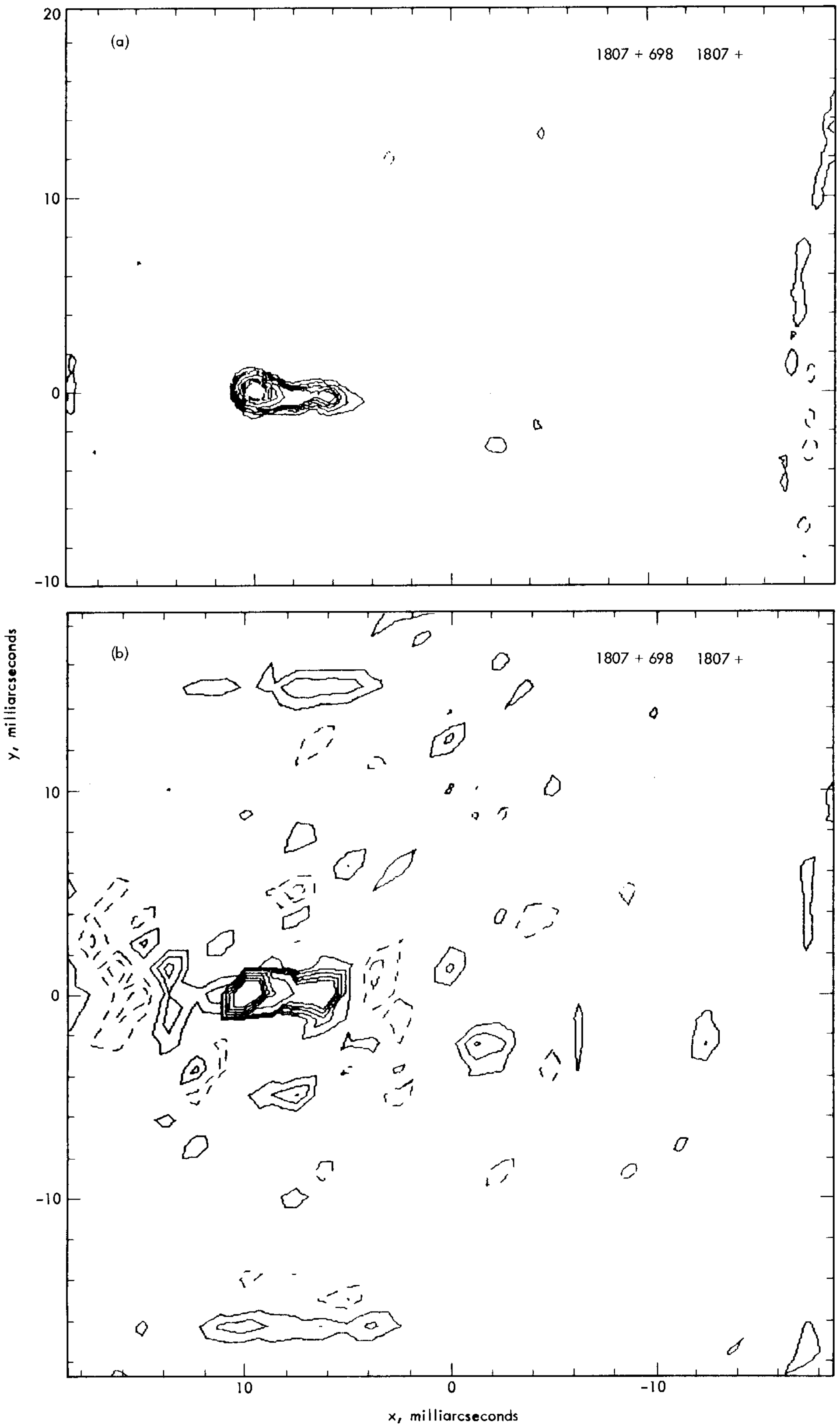

Fig. 4. Output of CLEAN: (a) 128 × 128 resolution; (b) 32 × 32 resolution

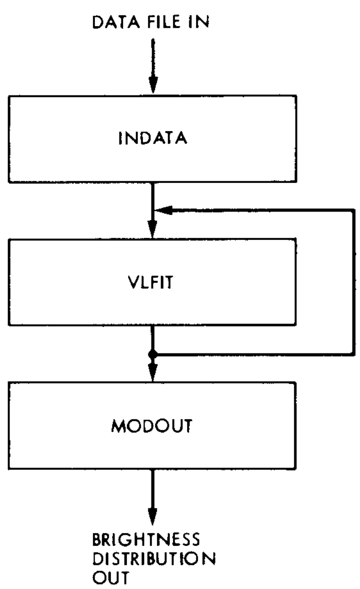

Fig. 5. Relation of the programs

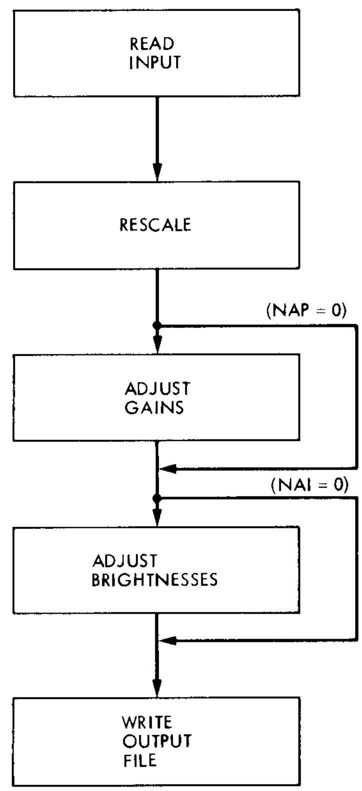

Fig. 6. The program VLFIT

Appendix A
Derivation of the Cost Function

In determining the brightness distribution of a radio source, preliminary reduction of VLBI data yields corrupted values of the visibility function

$$V(u,v) = \iint I(x,y) \exp\,[2\pi i(ux + vy)]\,dx dy \quad \text{(A1)}$$

[Refs. A-1 and A-2]. Here $I(x,y)$ is the brightness distribution, expressed as a function of rectangular coordinates x,y in the tangent plane to the celestial sphere at the source location. The visibility function $V(u,v)$ is its two-dimensional Fourier transform.

The core of the problem of finding the brightness distribution is the inversion of this transform. This is complicated by the fact that only values of $V(u,v)$ on a restricted, irregularly spaced set of points (u,v) are contained in the data.

Suppose there are NS antennas. Then the quantities u_p, v_p are associated with the p^{th} antenna, $1 \leqslant p \leqslant NS$, as follows. Let X_p be the vector from the center of the earth to the p^{th} antenna, and λ be the wavelength of the radiation received. Let K be the factor by which values of x,y must be multiplied to convert to radians (typically, x and y are measured in milliarcseconds, and $K = \pi \cdot 10^{-6}/648$). Then u_p, v_p are the components of KX_p/λ parallel to the x and y axes.

The only values of $V(u,v)$ which enter into the data are

$$V_{pq} = V(u_p - u_q, v_p - v_q), \qquad 1 \leqslant p,q \leqslant NS.$$

These quantities are functions of time, since u_p, v_p change due to the rotation of the earth. The data are averaged over time intervals sufficiently small that there is no significant variation of u_p,v_p over an interval. This reduces the set of visibilities to a finite set of values $V_{pq}^{(n)}$, where n is the index of the time interval. For a given value of n, some pairs p,q may not occur, because the source was not visible from certain antennas at that time, or because some data were lost.

To determine $I(x,y)$ numerically, it is approximated by a series of delta functions on a rectangular grid in a restricted region of the plane,

$$I(x,y) = \sum_{m=1}^{MF} I_m\, \delta(x - x_m)\, \delta(y - y_m).$$

Then equation (A1) is reduced to

$$V_{pq}^{(n)} = \sum_{m=1}^{MF} C_{pq,m}^{(n)}\, I_m\,, \quad \text{(A2)}$$

where

$$C_{pq,m}^{(n)} = \exp[2\pi i\{(u_p^{(n)} - u_q^{(n)})\,x_m + (v_p^{(n)} - v_q^{(n)})\,y_n\}]\,.$$

Unfortunately, there are additional complications in the problem. Each antenna has a gain, M_p, and a phase offset, ϕ_p, which are unknown functions of time. The effects of noise must also be taken into consideration. If $A_p^{(n)} = M_p \exp(i\phi_p)$, then the quantities actually given by the data are

$$E_{pq}^{(n)} = A_p^{(n)}\, \overline{A_q^{(n)}}\, V_{pq}^{(n)} + n_{pq}^{(n)} \quad \text{(A3)}$$

where the bar denotes a complex conjugate, and $n_{pq}^{(n)}$ is the contribution of the noise [Ref. 3).

Our problem is to determine the values I_m, given the quantities $E_{pq}^{(n)}$. If we start with assumed values $\hat{A}_p^{(n)}$, $\hat{I}_m$, we can construct

$$\hat{V}_{pq}^{(n)} = \sum_{m=1}^{MF} C_{pq,m}^{(n)}\, \hat{I}_m\,,$$

and

$$F_{pq}^{(n)} = \hat{A}_p^{(n)}\, \overline{\hat{A}_q^{(n)}}\, \hat{V}_{pq}^{(n)}\,. \quad \text{(A4)}$$

The noise terms $n_{pq}^{(n)}$ in (A3) are assumed to be unknown, so a criterion for the goodness of the assumed values is how close the $F_{pq}^{(n)}$ are to the $E_{pq}^{(n)}$. The averaging process which furnishes the data $E_{pq}^{(n)}$ also provides estimates, $\sigma_{pq}^{(n)}$, of the standard deviation of the noise. Thus we are led to consider the cost function

$$Q = \sum_n \sum_{p,q} \left| F_{pq}^{(n)} - E_{pq}^{(n)} \right|^2 \Big/ {\sigma_{pq}^{(n)}}^2\,. \quad \text{(A5)}$$

Minimization of Q gives the least-squares fit of the $F_{pq}^{(n)}$ to the $E_{pq}^{(n)}$.

References

A-1. Pachoczyk, A. G., *Radio Astrophysics*, Freeman, San Francisco (1970).

A-2. Fomalont, E. B., and Wright, M. C. H., in *Galactic and Extragalactic Radio Astronomy* (eds. Verschunt, G. L. and Kellermann, K. I.), Springer, New York (1974).

Appendix B
Iterative Minimization of the Cost Function

As stated in Section II, this minimization is accomplished by iterating two steps.

I. Variation of $\hat{I}_m$

Let $z_m = \hat{I}_m$. Then, in its dependence on the z_m, Q is a positive definite quadratic form

$$Q = \sum_{m,k=1}^{MF} B_{mk} z_m z_k - 2 \sum_{m=1}^{MF} D_m z_m + R , \quad \text{(B1)}$$

where

$$B_{mk} = \text{Real} \left\{ \sum_n \sum_{p,q} \left| \hat{A}_p^{(n)} \hat{A}_q^{(n)} \right|^2 C_{pq,m}^{(n)} \overline{C_{pq,k}^{(n)}} \Big/ \sigma_{pq}^{(n)^2} \right\},$$

$$D_m = \text{Real} \left\{ \sum_n \sum_{p,q} A_p^{(n)} \overline{A_q^{(n)}} \; C_{pq,m}^{(n)} \overline{E_{pq}^{(n)}} \Big/ \sigma_{pq}^{(n)^2} \right\},$$

and

$$R = \sum_n \sum_{p,q} \left| E_{pq}^{(n)} \right|^2 \Big/ \sigma_{pq}^{(n)^2} .$$

This quadratic form is minimized by the values of the z_m at which all the partial derivatives with respect to the z_m are zero:

$$\sum_{k=1}^{MF} B_{mk} z_k = D_m , \quad 1 \leqslant m \leqslant MF \quad \text{(B2)}$$

Direct solution of this system is not practical because of the large number of unknowns. An iterative method which converges to the solution is the *Gauss-Seidel* method [Ref. B-1], which consists of the following: Start with any assumed values of the unknowns. For $m = 1$ to MF, make the replacement

$$z_m \leftarrow \left(D_m - \sum_{k \neq m} B_{mk} z_k \right) \Big/ B_{mm} . \quad \text{(B3)}$$

When applied repeatedly, this procedure is easily shown to converge to the minimum of Q (for fixed $\hat{A}_p^{(n)}$). The step (B3) gives the minimum of Q when only the one unknown z_m is allowed to vary.

II. Variation of the $A_p^{(n)}$

Here each of the subsums of (A5),

$$Q_n = \sum_{p,q} \left| F_{pq}^{(n)} - E_{pq}^{(n)} \right|^2 \Big/ \sigma_{pq}^{(n)^2} ,$$

involves a separate set of variables, so they can be minimized separately. Let $a_p = \hat{A}_p^{(n)}$. Then

$$Q_n = \sum_{p,q} \left(G_{pq} \left| a_p a_q \right|^2 - 2 H_{pq} a_p \overline{a_q} \right) + R_n , \quad \text{(B4)}$$

where

$$G_{pq} = \left| \hat{V}_{pq}^{(n)} \right|^2 \Big/ \sigma_{pq}^{(n)^2} ,$$

$$H_{pq} = \hat{V}_{pq}^{(n)} \overline{E_{pq}^{(n)}} \Big/ \sigma_{pq}^{(n)^2} ,$$

and

$$R_n = \sum_{p,q} \left| E_{pq}^{(n)} \right|^2 \Big/ \sigma_{pq}^{(n)^2} .$$

This function is minimized by the gradient method [Ref. B-1]. Let

$$b_p = - \frac{\partial Q_n}{\partial \overline{a_p}} = 2 \sum_q - \left(G_{pq} |a_q|^2 \; a_p + H_{qp} a_q \right).$$

For $\nu > 0$, the quantities $a_p + \nu b_p$ are directed away from a_p along the negative gradient of Q_n (in the $2 \cdot NS$-dimensional space of the real and imaginary parts of the a_p's). The iterative step here is to replace a_p by $a_p + \nu b_p$ in the expression (B4) for Q_n, then choose ν to minimize the resulting fourth degree polynomial in ν. Using the minimizing value, $a_p + \nu b_p$ gives the new value of a_p.

In Step (2), as in (1), convergence is guaranteed since each step decreases Q. More generally, the steps of (1) and (2) can be intermingled in any systematic iterative procedure which varies each unknown infinitely often, and Q will approach the minimum. It is advantageous, however, to carry out Steps (1) and (2) separately to the point where the minimums of the separated problems are approached, because this reduces the amount of time spent in computing the coefficients in the formulas (B1) and (B4).

Reference

B-1. Burden, R. L., Faires, J. D., and Reynolds, A. C., *Numerical Analysis*, Prindle, Weber & Schmidt, Boston (1978).

Appendix C

Description of the Computer Programs

The method described here was implemented on the VAX computer PHOBOS in the Caltech Astronomy department. A series of three programs is used (see Fig. 5):

(1) The program INDATA reads the data from an existing data file in the MERGE format currently used at Caltech. This data is transferred into new files, with some conversion, and some of the quantities to be used in the minimization process are pre-computed.

(2) The program VLFIT carries out the minimization procedure of Section II.

(3) The program MODOUT reads the file output by VLFIT and builds a new file MOD.DAT which can be used by the Caltech program MODPLOT to draw a contour map of the source.

I. Program INDATA

This program reads an input file prepared by other existing programs. The first section reads a collection of header records, saving some of the information and converting some to a more convenient form. This information includes the astronomical location of the source (right ascension and declination), the starting time of the data, and names and locations of all the antennas. The time origin is chosen to be the start of the day on which the data begins, in Greenwich mean time. The right ascension is converted to *PLOØ*, the longitude of the source at time Ø. The vector from the earth's center to the p^{th} antenna, (STX(p), STY(p), STZ(p)), is rescaled so that its components in the xy-plane at the source are the quantities $2\pi u_p$, $2\pi v_p$ of Appendix A.

Next the data section of the file is read. The data points are stored sequentially in an array of length *IP*. The antenna pair (I,J) is denoted by one index $L = L(I,J)$ which is related to I and J by the arrays *IIB, JIB* set up in the program. We build the arrays

$$T(N), \quad N{=}1, \ldots, NT$$

and

$$E(K), SGM(K), LP(K), KT(K), \quad K{=}1, \ldots, IP\,,$$

where $T(N)$ is the time of the n^{th} time interval in minutes, and

$$E(K) = E_{pq}^{(N)}\,,$$

$$SGM(K) = \sigma_{pq}^{(N)^2}\,,$$

$$LP(K) = L(p,q)\,,$$

with

$$KT(K) = N, \text{ in the } n^{\text{th}} \text{ time interval.}$$

Data which contain no useful information are eliminated.

Next the program reads two lines of input data which specify parameters of the map. These are:

(1) NX,NY: Number of grid points in the x and y directions.

(2) XL, YL: Half-width of the map in x and y directions, in milliarcseconds.

The total number of real unknowns is now known to be $2NS{\cdot}NT + MF$ (where $MF = NX{\cdot}NY$). The number of real conditions in the data is $2{\cdot}IP$. The ratio $2IP/(2NS{\cdot}NT{+}MF)$, the redundancy, is evaluated here. If this value is less than 1, there are not enough conditions to determine the unknowns. If this happens, we may still get a good map, but there are other solutions to the minimum problem.

The next section of the program computes the quantities $2\pi u_p$, $2\pi v_p$ for each time interval, and forms the auxiliary complex arrays $CL\emptyset(p,N)$, $CLX(p,N)$, $CLY(p,N)$. These are used in VLFIT to generate $C_{pq,m}^{(n)}$ as follows: The x and y values at the grid points are

$$x_M = x_o + (I-1)DX, \quad 1 \leqslant I \leqslant NX,$$

$$y_M = y_o + (J-1)DY, \quad 1 \leqslant J \leqslant NY,$$

where $M = J + (I-1)NY$. If we put

$$CL(p,M,N) = CL\emptyset(p,N) \cdot CLX(p,N)^{I-1} \cdot CLY(p,N)^{J-1},$$

then

$$C_{pq,M}^{(N)} = CL(p,M,N) \cdot \overline{CL(q,M,N)}.$$

Finally, the brightness array $AI(M)$ and the complex gain array $AP(I,N)$ are initialized with values 1, and all the arrays needed by VLFIT are written in files.

II. Program VLFIT

This program reads the files prepared by INDATA, and two other lines of input: (i) NAP and NAI, and (ii) DSC (see Fig. 6). The gains are not adjusted if NAP = 0. The brightnesses are only rescaled if NAI = 0. DSC is a parameter used in image enhancement. For DSC > 0, the peaks of the brightness distribution are sharpened twice during the iterative solution for the brightnesses $AI(M)$.

First the brightnesses are rescaled so that for the given gains and shape of the brightness distribution, the fit to the data is as good as possible.

Next (if NAP≠0) step (2) of Appendix B is applied, with twelve iterations in each time interval.

Next (if NAI≠0) step (1) is applied. The values of $AI(M)$, M=1 to MF, are adjusted fifty times. After the 6th time and the 26th time, image enhancement is applied, by replacing each $AI(M)$ by

$$AI(M) - DSC \cdot (\max (AI) - AI(M))$$

and then replacing any negative values by zero.

Finally, an output file of values of gains and brightnesses is written, in the same format as the input file. This allows the application of VLFIT any number of times, using each output file as input the next time.

III. Program MODOUT

This program converts the output of VLFIT into a form useable by the program MODPLOT. The brightness distribution is characterized by a series of components, in this case delta-functions at the grid points, whose position is given in polar coordinates. These are listed in order of decreasing strength, terminating after at most MAXC components, where MAXC is a number read as input by the program.